

## **REMARKS**

Applicants have amended Claim 1 to explicitly indicate the result of the comparison between the desired attributes and respective desired attribute values and the attributes and respective values for the entire population is a target group.

The Examiner has rejected Claims 1-25 under 35 U.S.C. 103(a) as being unpatentable over *Piatetsky-Shapiro* (Shapiro), "Discovery, Analysis, and Presentation of Strong Rules", in "Knowledge Discovery in Database", AAAI/MIT Press, 1991, in view of *Simoudis, et al.* (U.S. Patent No. 5,692,107) and *Dash, et al.*, "Dimensionality Reduction of Unsupervised Data", Proceedings, Ninth IEEE International Conference on Tools with Artificial Intelligence, Nov. 1997. Applicants respectfully traverses such rejection.

The Examiner has cited *Shapiro* as teaching comparing attribute values having a desired attribute value to attribute values for all samples. Applicants contend *Shapiro* only teaches an algorithm for discovering exact rules in databases (see Introduction and the title heading of 13.5). The claimed invention performs the comparison on attributes and respective values to obtain a target group and not to discover exact or strong rules in a database. Applicants urge the Examiner to consider the difference between a database rule and a target group.

The following example serves to help bring to light the difference between the cited reference and the claimed invention, and to demonstrate the difference between a target group and a database rule. Consider the following sample population.

AGE	GENDER	OWNS A CAR	CAN LEGALLY DRIVE
10	MALE	NO	NO
24	FEMALE	YES	YES
40	FEMALE	NO	YES
22	MALE	YES	YES

The teachings of *Shapiro* allow one to determine rules between the different attributes of the sample population. Specifically, the algorithm in *Shapiro* results in the following set of rules:

If (AGE >= 24)	→ CAN LEGALLY DRIVE(YES)
If (AGE = 10)	→ CAN LEGALLY DRIVE (NO)
If (OWNS A CAR(YES))	→ CAN LEGALLY DRIVE(YES)

The purpose of using the algorithm in *Shapiro* is to discovery information not explicitly in the data. For example, the sample population does not indicate that one who owns a car can legally drive. Again, the results of the teachings of *Shapiro* are merely database rules. These rules are not the same as a target group in the claimed invention. Specifically, the information in the derived rules does not contain all of the data in a target population.

Using the same sample population, a target population as in the present invention is illustrated for people who own a car.

AGE	GENDER	OWNS A CAR	CAN LEGALLY DRIVE
24	FEMALE	YES	YES
22	MALE	YES	YES

The target population is simply the samples which have a desired attribute value. As the example illustrates, a database rule and a target population differ. The derived rules would not allow one skilled in the art to practice the present invention as the derived rules do not contain all of the necessary data. Consequently, Applicants contend *Shapiro*, when combined with the remaining cited references, does not teach nor suggest the present invention.

Additionally, Applicants contend the summary found in *Shapiro* to derive rules is not a target population. The summary in *Shapiro* does not preserve all of the field values and their relation to one another. At most, the summary in *Shapiro* keeps a partial count of the occurrence of a value. Such

a summary is unsuitable for generating a predictive model.

In rejecting Claim 13, the Examiner notes that *Shapiro* teaches the algorithm may be extended to handle multi-field conditions. Applicants believe the Examiner is misinterpreting the reference as in the discussion above. Further, Applicants contend the reference teaches the discovery of rules with complex antecedents (such as IF (A and B) THEN C). *Shapiro* does not teach comparing multiple attributes of a target population with attribute values of the entire sample population. The *Shapiro* reference is void of any mention of a target group. Therefore, Applicants contend the claimed invention is patentable over the cited references.

In rejecting Claims 14-19, the Examiner, referencing *Shapiro*, contends "the full data set can be estimated by sample-derived rules." Applicants urge the Examiner to realize that an *estimated* data set that may be derived from rules does not contain desirable information for generating accurate predictive models. Applicants refer to the above discussion to explain why an estimated data set is insufficient. Applicants urge the Examiner to consider the estimated data set will not preserve samples that may indicate a non-exact relationship between rules. For example, if two rules exist  $A \rightarrow B$  and  $C \rightarrow D$ , the samples A, B, C, D and A, B and C, D will not be preserved as no exact rule exists to create them in the estimate data set. Further, the Examiner has not provided a cite in *Shapiro* supporting the proposition that the full data set can be estimated by sample-derived rules. Additionally, Applicants find no reference to generating an estimated full data set is found in *Shapiro*.

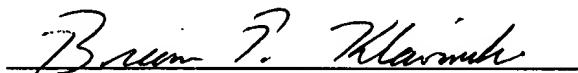
For the above reasons, Applicants contend the references cited by the Examiner, in combination or individually, do not teach the present invention. Combining *Dash*, which the Examiner cites as teaching the use of an entropy measure, and *Shapiro*, results in a set of exact rules and an entropy measure that has no application. Additionally, combining *Shapiro*, with *Simoudis*, results in no additional benefit from *Shapiro* alone as *Shapiro* discovers exact rules and does not preserve the data. *Simoudis* would have no practical application with *Shapiro* as no additional information may be mined from the exact rules resulting from *Shapiro*.

## CONCLUSION

Claims 1-25 are pending. Applicant believes independent claims 1, 13, 14, 20, 21 and 25 are distinguished over the cited references.

No extension of time or fee is believed to be necessary; however, in the event that any extension of time or fee is required for the prosecution of this application, please charge it against the IBM Corporation Deposit Account No. **09-0447**.

Respectfully submitted,



Brian P. Klawinski  
Registration No. 51,087  
BRACEWELL & PATTERSON, L.L.P.  
Suite 350, Lakewood on the Park  
7600B North Capital of Texas Highway  
Austin, Texas 78731  
(512) 542-2100

ATTORNEY FOR APPLICANT